基于深度卷积自编码神经网络的手写数字识别研究 *

曾文献，孟庆林，郭兆坤

(河北经贸大学 信息技术学院，石家庄 050061)

摘　要：手写数字识别在各种应用环境中对准确率的要求极高，传统机器学习方法由于训练样本单一，易在实际应用中识别错误。针对在提高不同笔体下的手写识别准确率进行了研究，将深度卷积神经网络与自动编码器相结合，设计卷积自编码器网络层数，形成深度卷积自编码神经网络。首先采用双线性插值方法分别对 MNIST 数据集与一万幅自制中国大学生手写数字图片进行图像预处理，然后先使用单一 MNIST 数据集对深度卷积自编码神经网络进行训练与测试，最后使用 MNIST 与自制数据集中 5000 幅混合，再次训练该网络，对另外 5000 幅进行测试。实验数据表明，所提深度卷积自编码神经网络在 MNIST 测试集正确率达到 99.37%，相比于前人有效提高准确率;且 5000 幅自制数据集模型测试正确率达 99.33%，表明该算法实用性较强，在不同笔体数字上得到较高识别准确率，模型准确有效。

关键词：卷积自编码神经网络；双线性插值；手写数字识别；深度学习

中图分类号：TP391.41　　doi: 10.19734/j.issn.1001-3695.2018.09.0774

Research on handwritten digit recognition based on deep convolution self-coded neural network

Zeng Wenxian, Meng Qinglin, Guo Zhaokun

(*School of Information Technology*, *Hebei University of Economics & Business*, *Shijiazhuang* 050061, *China*)

Abstract: Handwritten digit recognition requires extremely high accuracy in various application environments. Traditional machine learning methods are easy to identify errors in migration applications due to the single training sample. The article studied the accuracy of handwriting recognition under different fonts, combined a deep convolutional neural network with an automatic encoder to design a convolutional self-encoder network layer to form a deep convolutional self-encoding neural network. Firstly, this paper used the bilinear interpolation method to preprocess the MNIST dataset and 10, 000 self-made Chinese college students' handwritten digital images. Then, this paper used a single MNIST dataset to train and test the deep conlutional self-encoding neural network. Finally, this paper mixed MNIST with 5, 000 self-made data sets, trained the network again, and tested another 5, 000. The experimental data showed that the correct rate of the deep convolutional self-encoding neural network in the MNIST test set reached 99.37%, which was more effective than the predecessors. And the correct rate of 5, 000 self-made dataset model tests was 99.33%. The result indicates that the algorithm has strong migration and application ability, and it has higher recognition accuracy on different pen Figures. The model is accurate and effective.

Key words: convolutional self-encoding neural network; bilinear interpolation; handwritten digit recognition; deep learning

0　引言

手写数字识别在计算机视觉领域占有非常重要的地位。它利用现有计算机技术及摄像头等设备，对日常生活中手写阿拉伯数字进行辨识，在财务处理、金融管理、税务管理等等领域应用广泛。然而由于人工手写数字笔体不一，使得计算机进行智能识别时准确率较低。例如在日常生活中进行银行支票处理或邮政编码识别时，极其微小的错误将会导致巨大的损失。因此，手写数字识别最重要的任务是提高识别准确率。

深度卷积神经网络在手写数字识别中应用极为广泛，且其性能也较为优良。但神经网络取决于训练数据集，当训练样本特征覆盖范围较大时，所训练网络的准确性将相对较高。但在神经网络模型建立中，若采用特征覆盖性较小的数据集进行训练测试，所训练的神经网络模型将会过度拟合此数据集，当使用此神经网络识别新的手写数字图片时，易出现识别错误。此外，卷积神经网络池化层会丢失较多的信息，从而降低了空间分辨率，这就导致对于输入的微小变化，其输出几乎是不变的，容易降低模型准确性。因此找到合适的算法以及特征覆盖范围较大的手写数字训练数据成为识别的关键。

自动编码器由编码器和解码器两部分组成，能够尽可能复现输入特征。作为一种无监督学习的非线性特征提取方法，其输出与输入具有相同的维度，隐藏层则被用来进行原始数据的特征表示或编码[1]。因其无监督特征提取方法的特点，其虽可快速提取特征，但特征提取准确性与还原性却无法保证。

本文将卷积神经网络与自动编码器结合，形成深度卷积自编码神经网络，根据手写数字图像特点，设置卷积核参数，以此进行特征提取。所提取特征不光具有自动编码器输出特征与输入相同、无监督快速提取的优点，还拥有卷积神经网络权值共享的优点。较大程度解决了卷积神经网络池化层丢

收稿日期：2018-09-14；**修回日期**：2018-10-19　**基金项目**：河北省科技计划项目（17450112D）

作者简介：曾文献（1971-），男，陕西旬阳人，教授，硕士，主要研究方向为深度学习、机器视觉等；孟庆林（1994-），男（通信作者），硕士研究生，主要研究方向为深度学习、机器视觉（2583813020@qq.com）；郭兆坤（1993-），男，硕士研究生，主要研究方向为深度学习、机器视觉.

失信息与自动编码器特征提取准确率低的缺点。

采用 MNIST 数据集作为训练与测试集，MNIST 是目前手写阿拉伯数字的最大数据集，特征覆盖范围广，可在同一个数据集上最大限度的解决卷积神经网络识别范围较窄的问题。同时，将 10000 幅自制中国大学生手写数字数据集分为各 5000 幅，选出 5000 幅与 MNIST 中训练集再次训练网络，用另外 5000 幅测试该网络，以佐证模型实用性能。经实验验证，MNIST 数据集上的识别准确率相对于前人方法有一定提升。此外，为改善像素特征并扩大数据集中不同图像数字特征的差别，故在图像预处理阶段使用了插值图像增强算法，将图像重新调整，使特征提取更为精准。

1 已有研究及本文相关算法

1.1 已有研究

前人已有多种方法实现手写数字识别，主要包括机器学习、深度学习、模糊控制等领域。结果表明，深度学习领域算法已超越大多数用于分类的算法，文献[2]提出的使用卷积神经网络与自动编码器对孟加拉数字数据集进行识别，其准确率到达 98.55%；文献[3]提出的“融合卷积神经网络”将 SN 模型与 B-CNN 模型融合，在 MNIST 上测试正确率达到 99.10%；文献[4]提出的基于雅克比稀疏自动编码机的模型，在 MNIST 上测试正确率达到了 94.8%；文献[5]使用“Dropconnect”方法实现了 MNIST 测试正确率 99.79%，这是目前此数据集上的最高正确率；文献[6]使用了局部二进制特征提取方法与 K-近邻分类算法，达到了 89.81%的正确率；文献[7]提出的无监督学习应用对象特征层次结构是一种无监督方式 MNIST 测试方法，其准确率达到了 98.32%；文献[8]建立的深度有监督学习网络，相比于其他文献所提取的特征，可将错误降低 50%；文献[9]将隐马尔可夫模型应用于手写数字识别，在银行票据 OCR 的应用中使得票据拒识别率降低了 3%；文献[10]使用了基于轮廓特征的 HMM 手写数字识别，达到了 92.2%的正确率；文献[11]使用的图像增强方法，配合卷积神经网络，在 MNIST 数据集上达到 99.56%的准确率。

1.2 所用基础方法

1.2.1 卷积神经网络（CNN）

卷积神经网络是一种特殊的可训练权重和偏置的网络，经典卷积神经网络图如图 1 所示。

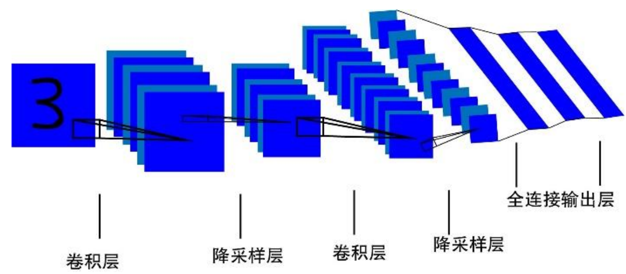


图 1 经典卷积神经网络图

Fig. 1 Classic convolutional neural network diagram

其与人工神经网络不同点在于局部感受域与权值共享。但当普通深度卷积神经网络层级加深，目标函数优化则将变为非凸优化问题，模型识别准确率将极度依赖初值的选择，若初值选择恰当，可使优化函数稳定收敛；若初值选择不当，模型将会出现欠拟合情况，在可行域上出现大量极值点。

1.2.2 自动编码器(autoencoder)

自动编码器是深度学习中的一个基础概念。可以使用自身的高阶特征进行自我编码。自动编码器其实也是一种神经网络，故其思想来源于人脑对外界认知的处理过程，模拟人视觉系统对外界事物的处理，在处理时，信号从一个神经元传递到下一个神经元，不同神经元的特征不断被抽象，相同神经元的特征被强化。其最终将使用高阶特征重构自己，而非复制像素点。

自动编码器是指保持输入与输出尽可能一致（通过信息损失来判定）的情形下，实现无监督方式下的隐层特征提取与参数学习[12]。目标是让神经网络的输出能和原始输入一致，相当于在特征空间上学习一个恒等式 $y=x$ 。将原始图像作为输入，对图像进行编解码，使提取到的特征保持输入与输出接近一致。自动编码器结构图如图 2 所示。

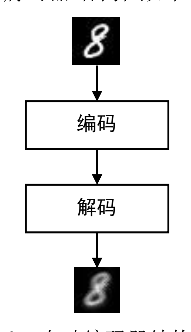


图 2 自动编码器结构图

Fig. 2 Automatic encoder structure

1.2.3 图像增强技术

在利用深度学习进行图像识别时，由于需要进行高度抽象的特征提取，这就需要图像特征高度有效，而训练集图像不一，缺乏泛化性与特征有效性。因此在进行模型训练之前，首先要对图像采取一系列的预处理，图像增强技术是预处理阶段的重要过程，可以提高特征有效性，加强模型泛化能力（防止过拟合）。

本文采取的图像增强技术是双线性插值增强，分别对图片 x 与 y 两个方向上的像素做插值，使图像进行像素尺寸缩放。达到图像增强效果。

2 核心算法建立

本文模型建立流程包括图像预处理方法；卷积自动编码器模型建立特征提取方法；人工神经网络模型建立进行训练与预测方法。本文手写数字识别方法流程如图 3 所示。

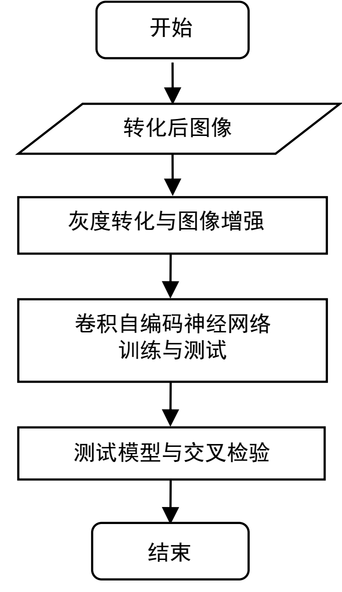


图 3 手写数字识别方法流程

Fig. 3 Flow chart of handwritten digit recognition method

2.1 图像预处理

图像预处理阶段分两步进行，第 1 阶段对 MNIST 数据集 28*28 像素图像进行灰度处理，将其转换为灰度图像，然后对灰度图像进行二值化处理，去除噪声；第 2 阶段进行图像增强，采用双线性插值增强方法对图像进行像素缩放，达到图像增强目的。

2.1.1 灰度转换

本文所使用的 MNIST 数据集与 10000 幅自制数据集图像像素均为 28*28 的黑白图像。首先将其转换为灰度图像，这里用到 Python 中的 numpy 库 convert 函数进行灰度转换。

转换为灰度图像后对图像进行二值化处理，本文灰度图像转换为二值化图像阈值 θ 计算公式如下：

$$\theta=\sum_{i=1}^{Rs}\sum_{j=1}^{Cs}h_{ij} \tag{1}$$

其中：Rs 表示像素行数，Cs 表示像素列数，h_{ij} 表示像素亮度值。根据此阈值转换为二进制图像方法为

$$K_{ij}=\begin{cases}1, if\ h_{ij}>\theta\\0, if\ h_{ij}<\theta\end{cases} \tag{2}$$

对于本文使用的 MNIST 与自制数据集，Rs=Cs=28。

2.1.2 图像增强

本文预处理阶段图像增强目的是为实现图像尺寸缩放进而加强图像边缘特征。设原图像大小为 m*n，缩放后的大小为 a*b。则两幅图像的边长比为 m/a 与 n/b，在图像缩放的大多数情况下，边长比不为整数，故存储时选用浮点型。则目标图像的第(x, y)个像素点（第 x 行第 y 列）可通过边长比对应至原图像，则对应坐标为$(x*m/a, y*n/b)$，则根据边长比的取值，此值也为非整型，而非整型的坐标无法在图像上表达。

采取双线性插值方法进行像素转换，可有效避免坐标是浮点数据的问题。若所求得原图像的对应坐标为(x*m/a, y*n/b)，则双线性插值算法首先寻找离其最近且紧邻的 4 矩形像素点，计算此像素点与周围 4 像素点的双线性插值。根据插值意义，在离散点 p_i 处给定被逼近函数 f(x)的值 f(pi)，则逼近函数 $y(p_i')$满足：

$$\max|f(p_i)-y(p_i')|=0 \tag{3}$$

执行插值操作后，可使逼近误差接近 0。即目标图像像素点与原始像素点逼近误差接近 0。减少图像增强过程中的图像失真。

已知目标图像上的第(x, y)个像素点 $P(x,y)$，则设其在原图像上的 4 个矩形相邻像素点，分别为 $P_{11}=(x_1, y_1)$、$P_{12}=(x_1, y_2)$, $P_{21}=(x_2, y_1)$ 以及 $P_{22}=(x_2, y_2)$，则在增强过程中首先在 x 方向进行插值，计算公式如下：

$$y(Q_1)\approx\frac{x_2-x}{x_2-x_1}f(P_{11})+\frac{x-x_1}{x_2-x_1}f(P_{21})\ \text{where}\ Q_1=(x,y_1) \tag{4}$$

$$y(Q_2)\approx\frac{x_2-x}{x_2-x_1}f(P_{12})+\frac{x-x_1}{x_2-x_1}f(P_{22})\ \text{where}\ Q_1=(x,y_2) \tag{5}$$

其中：$f(P_{ij})$ 表示在原图像中第 i 行第 j 列位置的像素值，$y(Q_1)$ 表示在 (x,y_1) 线上的 x 方向插值得到的像素值，其为插值过程中过渡值。

在 y 方向上插值为

$$F(R)\approx\frac{y_2-y}{y_2-y_1}y(Q_1)+\frac{y-y_1}{y_2-y_1}y(Q_2) \tag{6}$$

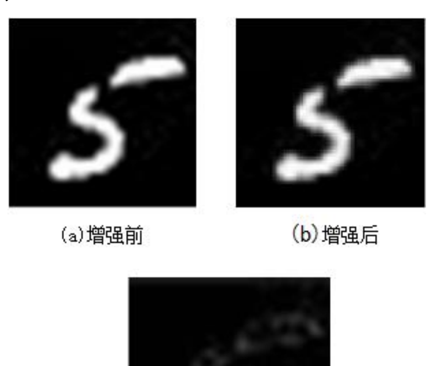
其中：$F(R)$ 表示在 y 方向上的插值结果，即在目标图像上的像素表达式。

将 x 方向与 y 方向插值综合，得到双线性插值结果：

$$F(x,y)\approx\frac{(x_2-x)(y_2-y)}{(x_2-x_1)(y_2-y_1)}f(P_{11})+\frac{(x-x_1)(y_2-y)}{(x_2-x_1)(y_2-y_1)}f(P_{21})+\frac{(x_2-x)(y-y_1)}{(x_2-x_1)(y_2-y_1)}f(P_{12})+\frac{(x-x_1)(y-y_1)}{(x_2-x_1)(y_2-y_1)}f(P_{22}) \tag{7}$$

其中：$F(x,y)$ 表示最终插值结果，即目标图像第 (x,y) 个像素点的像素值。

根据此像素插值计算方式，采用 MATLAB 对此算法进行编程实现，对 MNIST 数据集中某一张图片数字“5”进行缩放，先缩小到 25*25 像素尺寸，随后再增大到 28*28 像素尺寸，实现图像增强。图像增强后的图片及增强前后对比图，如图 4 所示。



(a)增强前　　(b)增强后

(c)增强前后对比

图 4　双线性插值增强效果对比

Fig. 4　Comparison of double interpolation enhancement

2.2 卷积自编码神经网络

卷积神经网络卷积层用来提取图像特征，采用卷积计算的方式，对像素与卷积核参数进行卷积运算。训练过程中，此层的卷积核会进行训练修正，以此降低损失。卷积计算公式如下所示：

$$conv(m,n)=(P\otimes Q)(m,n)=\sum_a\sum_b I(a,b)(m-a,n-b) \tag{8}$$

其中：$\otimes$ 表示卷积运算，此公式表示了卷积核 Q 在输入图像上在平面空间滑动，计算元素乘法求和产生输出，得到一个卷积特征图[13]。

在经过每个卷积层后，还会经过一池化层（降采样层），每次将原图像卷积后，都通过一个下采样的过程，来减小图像的规模，降低卷积后的过拟合。

卷积神经网络最大的优点在于局部感受域与权值共享。局部感受域通过卷积操作，把全连接变成局部连接，因为多层网络能够抽取高阶统计特性，即使网络为局部连接，由于格外的突触连接和额外的神经交互作用，也可以使网络获得全局关系。卷积核通过原图像，然后卷积核对原图像上符合卷积核大小的像素进行加权求和，每一次只是对符合卷积核的图像像素做卷积，，使全连接变成局部连接。

权值共享在不同的图像或者同一张图像共用一个卷积核，减少重复的卷积核。同一张图像当中可能会出现相同的特征，共享卷积核能够进一步减少权值参数。如果使用了权值共享，即共用同一个卷积核，那么将可以大大减少卷积核的数量，加快运算速度。

结合卷积神经网络与自动编码器的优点，利用自动编码器的不同神经元的特征不断被抽象，相同神经元的特征被强

化优点，克服卷积神经网络特征提取易出现欠拟合的问题，形成了卷积自动编码器。首先进行卷积编码操作，然后进行反卷积解码器解码操作。为实现特征全面有效提取与各层分工合作，适当改变编解码器中卷积核的大小与卷积层个数，实现相应特征提取功能。具体参数分别为：

a）**卷积编码器层**包含两个卷积层和两个池化层。

第 1 层卷积核(conv1)参数为 5*5*32,，采用全 0 填充。设置此层的目的是为提取图像的边缘范围特征，确定图片整体形状与图像位置。

第 1 层下采样层(Pool1)参数为 2*2，步长为 2,使用全 0 填充，采用平均池化方式。设置此层的目的是对边缘范围特征进行压缩，并采取平均池化保留背景。

第 2 层卷积核(conv2)参数为 5*5*64，使用全 0 填充。设置此层的目的是为提取图像的角特征与脊特征，以便更好识别图像手写数字中的边角与矩形位置，如数字“7”、“5”与数字“1”“9”中等，MNIST 中图像边角特征部分与脊特征部分如图 5 所示：

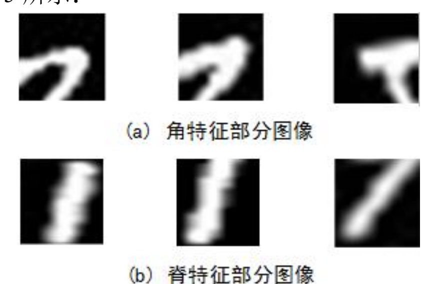


图 5 角特征与脊特征展示

Fig. 5　Angular feature and ridge feature display

b）**卷积解码器层**。卷积解码器层采用反卷积操作，对特征进行放大。将卷积中编码器层的结果逐层进行反卷积操作，最终使输出恢复至原图像大小，实现尺寸无损的特征提取。卷积解码器层包括 2 个反卷积层，2 个上采样层，输出为 28*28 尺寸特征。

第 1 层反卷积(Deconv1)参数为 2*2*64，步长为 2，不采用全 0 填充。卷积编码器层输出经过此层后，大小变为 14*14*64。

第 1 层上采样层（Upsampling1）参数为 3*3*32，采用 2 层全 0 填充。经过此层后，图像大小变为 14*14*32。设置 Deconv1 与 Upsampling1 的目的为将编码器层提取到的边角特征与脊特征进行特征映射，恢复一定尺寸特征。

第 2 层反卷积层(Deconv2)参数为 2*2*32，步长为 2，不采用全 0 填充。经过此层后输出变为 28*28*32。

第 2 层池化层(Upsampling2)参数为 3*3，步长为 1，采用 1 层全 0 填充，经过此层后，输出结果为 28*28*1。设置 Deconv2 和 Upsampling2 的目的为将卷积编码器中保留的背景映射到新的尺寸中，并使最终得到的特征图与原图尺寸一致。

本文卷积自动编码器模型如图 6 所示。此卷积自动编码器形成的卷积自编码神经网络图如图 7 所示。

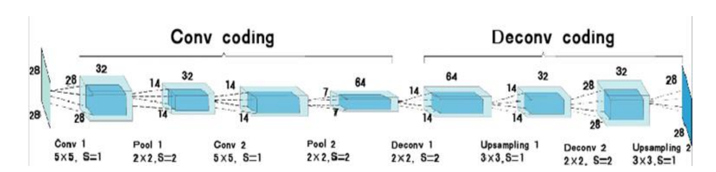


图 6　所设计卷积自动编码器

Fig. 6　Convolution autoencoder model designed

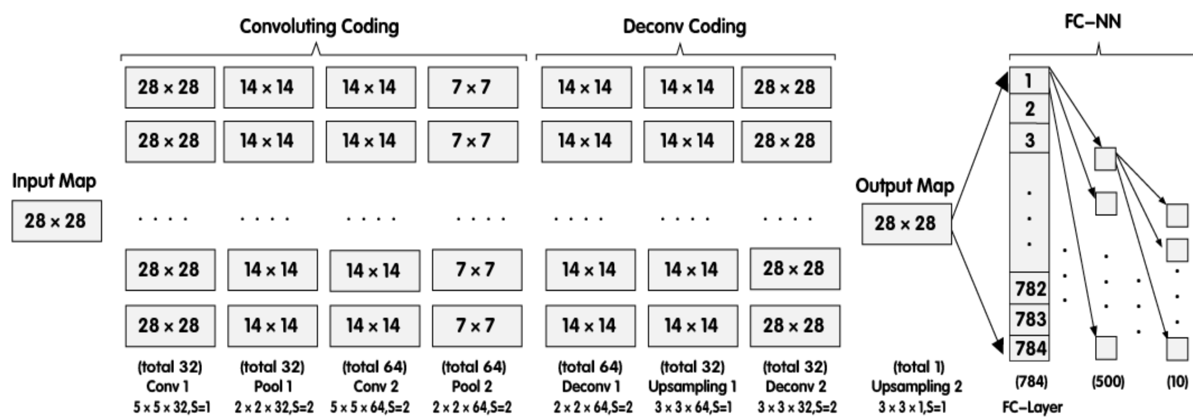


图 7　卷积自编码神经网络结构图

Fig. 7　Convolutional self-encoding neural network structure

卷积编码层通过卷积运算对图像进行编码，利用卷积核的参数共享机制对加快提取速度，且局部感受域可高度有效抽取图像特征。

全连接神经网络（FC-NN）输入层为 784 个神经元，隐含层为 500 个神经元，输出层为 10 个神经元。

网络所使用的激活函数为 ReLU 函数，代价函数采取 softmax 与交叉熵结合。首先构建神经元神经网络模型，其中每一层之间的连接采用可训练权重进行连接。权重矩阵 ω 初值采取截断的正态分布获取。且将 ω 加入 L_2 正则化，限制权重 ω 大小，使模型不能随意拟合训练数据随机噪声，即有效防止训练模型过拟合。

L_2 正则化公式如下：

$$R(\omega)=\|\omega\|_2^2=\sum_i \omega_i^2 \tag{9}$$

其中，$R(\omega)$ 的结果即为加入 L_2 正则化后的权重限制。

当输入特征经过第 1 层隐含层加权后，需将加权的神经元特征通过激活函数将特征保留并映射出去，解决实现非线性与线性的转换。这里本文选用的激活函数为 ReLU 函数，函数公式如下：

$$\mathrm{ReLU}(x_i)=\max(0,x_i) \tag{10}$$

由此函数的特点可知，当 x_i 逐渐变大，即随着训练次数不断增加，函数的变化率不变，故训练时，执行梯度下降算法，会较好的进行下去，且函数变化率也处在较好的范围内。

本文应用的全连接神经网络经过最后一层隐含层输出后，会经历一个原始输出层，随后经历 softmax 层，softmax 表达式如下：

$$\mathrm{softmax}(y)_i=y_{i'}=e^{y_i}\Big/\sum_{j=1}^{n}e^{y_j} \tag{11}$$

其中：y_i 表示神经网络的原始输出经过 softmax 层后，整体神经网络输出变成了一个符合概率分布的参数。

交叉熵表示式如下：

$$H(m,n))-\sum_{x} m(x)\log n(x) \tag{12}$$

*H(m,n)*表示 *m* 与 *n* 之间的概率分布距离。由 softmax 层得到概率分布参数后，将此参数与训练标签值进行交叉熵运算，运算后的结果即得到前向传播输出与概率分布之间的概率距离。训练数据的标签值就符合一个概率分布，即属于正确类别的概率为 1，属于不正确类别的概率为 0，实现前向传播第一次预测。随后进行训练，即反向传播，通过最小化损失函数，对权值 ω 进行微分运算，找到最小化损失函数的相应权值与偏置，卷积自编码神经网络训练完毕。

3　数据集与实验结果

本部分首先介绍使用的数据集，介绍所使用程序运行工具，再对实验结果与交叉检验结果进行展示分析。

3.1　数据集

MNIST 数据集:MNIST 是美国国家技术与标准研究所收集制作的阿拉伯手写数字数据集，由 250 个不同的人手写而来。共包含 70000 图片，其中 60000 张为训练数据集，10000 张为测试数据集。MNIST 数据集部分图片如图 8 所示。

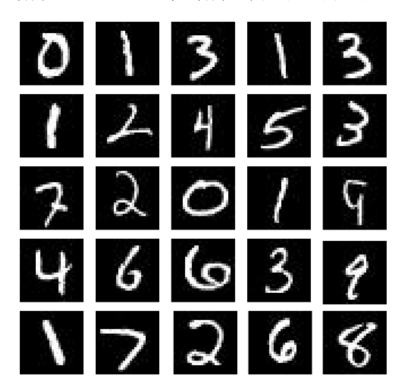


图 8　MNIST 数据集部分图片

Fig. 8　MNIST data set part of the picture

自制数据集：除 MNIST 数据集外，本文还用了 10000 幅自制手写数字数据集。其中数字 0-9 各 1000 张，由 70 名中国本科生与研究生手写而成。制作数据集，将像素转换为 28*28 像素，进行图像预处理，自制数据集图片如图 9 所示。

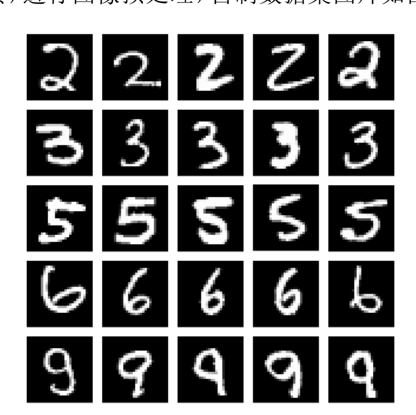


图 9　自制数据集部分图片

Fig. 9　Homemade dataset part of the picture

实验中，将 MNIST 训练集与自制数据集图片进行上文的预处理操作，灰度转换并进行图像增强，图像增强的过程中存在像素缩放，具体为：

a）MNIST 训练集。图片原始像素为 28*28，首先将其压缩至 25*25 像素，再放大至 28*28 像素。

b）自制数据集。图片原始像素为 28*28，首先将其压缩至 25*25，再放大至 28*28 像素将处理后的图片保存。

3.2　实验结果与对比分析

实验环境为 Windows7 系统，采用编程语言为 Python，神经网络学习框架采用 TensorFlow 框架。实验中，将上文提出的深度卷积自编码神经网络对 MNIST 预处理后的图片进行特征提取与训练。经过 40000 步训练后，损失趋近于稳定。将训练好的模型对 MNIST 训练集进行测试，测试准确率为 99.37%。如图 10 所示。

```
After 32001 training step(s),test accuracy = 0.9919
After 33001 training step(s),test accuracy = 0.99256
After 34001 training step(s),test accuracy = 0.9927
After 35001 training step(s),test accuracy = 0.9928
After 36001 training step(s),test accuracy = 0.9920
After 37001 training step(s),test accuracy = 0.99297
After 38001 training step(s),test accuracy = 0.9930
After 39001 training step(s),test accuracy = 0.9937

Process finished with exit code 0
```

图 10　MNIST 训练模型测试结果

Fig. 10　MNIST training model test results

将预处理后的一万幅中国大学生手写数字图片分为各 5000 幅图像，选其中 5000 幅与 MNIST 训练集一起组成新的训练集，再次训练神经网络，经过 40000 步训练。将训练好的模型对另外 5000 幅图像进行测试，佐证模型实用性，测试准确率为 99.33%。如图 11 所示。

```
After 31001 training step(s),test accuracy = 0.9913
After 32001 training step(s),test accuracy = 0.9925
After 33001 training step(s),test accuracy = 0.99268
After 34001 training step(s),test accuracy = 0.99287
After 35001 training step(s),test accuracy = 0.9927
After 36001 training step(s),test accuracy = 0.99281
After 37001 training step(s),test accuracy = 0.99293
After 38001 training step(s),test accuracy = 0.9931
After 39001 training step(s),test accuracy = 0.9933

Process finished with exit code 0
```

图 11　自制数据集训练模型测试结果

Fig. 11　Self-made data set training model test results

在 MNIST 上的模型训练过程损失率变化在 Tensorboard 可视化后，得到损失率变化曲线，如图 12 所示。

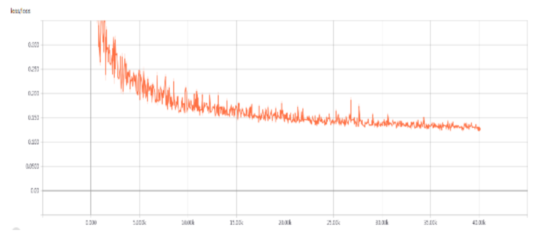


图 12　MNIST 上训练损失率变化

Fig. 12　Change in training loss rate on MNIST

在自制数据集上的损失率变化曲线，如图 13 所示。

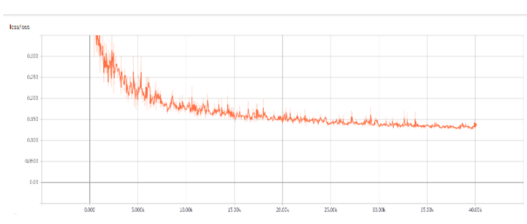


图 13　自制数据集上训练损失率变化

Fig. 13　Changes in training loss rate on the homemade data set

由损失率变化曲线，模型在训练 30000 步后，已基本接近稳定，模型稳定有效。

卷积自动编码器与神经网络结合方法与其他方法对比如表 1 所示。

表 1 算法 MNIST 测试准确率对比

Table 1 Comparison of algorithm MNIST test accuracy

分类算法	准确率
K-近邻	89.81%
基于轮廓的 HMM	92.20%
B-CNN	99.10%
卷积自编码神经网络	99.37%

4 结束语

本文介绍了一种手写数字的识别方法，首先使用双线性插值对图像进行预处理，然后利用卷积自动编码器与人工神经网络，在 MNIST 数据集训练并测试，MNIST 数据集识别正确率达到 99.37%。使用部分自制数据集与 MNIST 混合再次训练模型，此模型对剩余自制数据集进行测试，测试集识别正确率达到 99.33%。模型稳定，识别高效准确。

文中提出的方法虽未达到前人在 MNIST 数据集上的最高准确率，为解决手写数字识别问题提供了不错的思路。且该方法不只可应用在手写数字识别领域，还可应用在手写汉字识别与人脸识别领域，下一步将考虑在此方向进行实验探究。

参考文献：

[1] 张慧，王坤峰，王飞跃. 深度学习在目标视觉检测中的应用进展与展望 [J]. 自动化学报，2017, 43(8): 1289-1305. (Zhang Hui, Wang Kunfeng, Wang Feiyue. Advances and Perspectives on Applications of Deep Learning in Visual Object Detection [J]. Acta Automatica Sinica. 2017, 43(8): 1289-1305.)

[2] Shopon M, Mohammed N, Abedin M A. Bangla handwritten digit recognition using autoencoder and deep convolutional neural network [C]//Proc of International Workshop on Computational Intelligence. Piscataway,NJ:IEEE Press, 2017: 64-68.

[3] 陈玄，朱荣，王中元. 基于融合卷积神经网络模型的手写数字识别 [J]. 计算机工程，2017, 43 (11): 187-192. (Chen Xuan, Zhu Rong, Wang Zhongyuan. Handwritten digits recognition based on fused convolutional neural network model [J]. Computer Engineering, 2017, 43 (11): 187-192.)

[4] 王慧玲，宋威. 基于雅克比稀疏自动编码机的手写数字识别算法 [J]. 计算机应用研究, 2018, 35(12): 3816-3819, 3823. (Wang Huiling, Song Wei. JSAE: Jacobian regularized sparse auto-encoders on MNIST database [J]. . Application Research of Computers, 2018, 35(12): 3816-3819, 3823.)

[5] Wan L, Zeiler M D, Zhang S, *et al.* Regularization of neural networks using DropConnect [C]//Proc of International Conference on Machine Learning. 2013: 1058-1066.

[6] Ilmi N, Budi W T A, Nur R K. Handwriting digit recognition using local binary pattern variance and K-nearest neighbor classification [C]// Proc of International Conference on Information and Communication Technology. IEEE, 2016: 1-5.

[7] Marc, Huang F J, Boureau Y, *et al.* Unsupervised learning of invariant feature hierarchies with applications to object recognition [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition.2007: 1-8.

[8] 陈浩翔，蔡建明，刘铿然，等. 手写数字深度特征学习与识别 [J]. 计算机技术与发展，2016, 26 (7): 19-23. (Chen Haoxiang, Cai Jianming, Liu Kengran, *et al.* Deep learning and recognition of handwritten numeral features [J]. Computer Technology and Development. 2016, 26 (7): 19-23.)

[9] 刘刚，张洪刚，郭军. 用于脱机手写数字识别的隐马尔可夫模型 [J]. 计算机研究与发展，2003, 40(8): 1252-1257. (Liu Gang, Zhang Honggang, Guo Jun. Application of HMM in handwritten digit OCR [J]. Computer Technology and Development. 2003, 40 (8): 1252-1257.)

[10] 肖明，贾振红. 基于轮廓特征的 HMM 手写数字识别 [J]. 计算机工程与应用，2010, 46 (33): 172-174. (Xiao Ming, Jia Zhenhong. HMM handwritten digit recognition based on contour features [J]. Computer Engineering and Applications. 2010, 46 (33): 172-174.)

[11] Shopon M, Mohammed N, Abedin M A. Image augmentation by blocky artifact in deep convolutional neural network for handwritten digit recognition [C]//Proc of IEEE International Conference on Imaging, Vision & Pattern Recognition. IEEE, 2017: 1-6.

[12] Teow M Y W. A minimal convolutional neural network for handwritten digit recognition [C]//Proc of IEEE International Conference on System Engineering and Technology. IEEE, 2017: 171-176.

[13] 焦李成，赵进，等. 深度学习、优化与识别 [M]. 北京：清华大学出版社，2017:84-85. (Jiao Licheng, ZhaoJin, *et al.* Deep learning, optimization and recognition [M]. Beijing: Tsinghua University Press, 2017: 84-85.)